Research Article

Licheng Li*

Research on target feature extraction and location positioning with machine learning algorithm

Abstract: The accurate positioning of target is an important link in robot technology. Based on machine learning algorithm, this study firstly analyzed the location positioning principle of binocular vision of robot, then extracted features of the target using speeded-up robust features (SURF) method, positioned the location using Back Propagation Neural Networks (BPNN) method, and tested the method through experiments. The experimental results showed that the feature extraction of SURF method was fast, about 0.2 s, and was less affected by noise. It was found from the positioning results that the output position of the BPNN method was basically consistent with the actual position, and errors in X, Y and Z directions were very small, which could meet the positioning needs of the robot. The experimental results verify the effectiveness of machine learning method and provide some theoretical support for its further promotion and application in practice.

Keywords: intelligent robot, binocular vision, feature extraction, machine learning, speeded-up robust features, back propagation neural network

2020 Mathematics Subject Classification: 68T40

1 Introduction

With the development of technology, intelligent robots have been more and more widely used in people's work and life, including civil [1], industry [2], agriculture [3], medical [4], military [5], etc., especially in environments such as the universe and ocean that human beings can not match. Intelligent robots is an important embodiment of national science and technology level and industrial level. In order to achieve more intelligent service of robots, it is necessary to improve the control technology of robots, such as target positioning, path planning, etc. [6]. Taking industrial robot as an example, in the process of completing tasks, the robot needs to locate the target accurately to realize the following work such as recognition, detection, grasping and classification. Robot positioning has become a problem which is widely concerned by researchers [7]. Zhang *et al.* [8] designed a piecewise fitting monocular vision ranging method for the positioning of humanoid robots and found through experiments that the average error of the method was 1.7 mm, suggesting relatively accurate positioning. Luo *et al.* [9] studied the positioning of picking points of grape picking robot. Based on the improved clustering image segmentation algorithm, the positioning of picking points was carried out. Experiments were carried out in the environment of Opencv2.3.1 and Visual C + +. They found that the error ratio of the picking points obtained by the proposed method and the manually set picking points reached 88.33%, which could meet the position requirements of robot. Yan *et al.* [10] designed a positioning method based on passive radio frequency identification (RFID), carried out simulation experiments, and found that the absolute error of the method was smaller than 10.16 cm and the calculation time was short. Wu *et al.* [11]

*Corresponding Author: Licheng Li:** School of Information Science and Engineering, Hunan First Normal University, Changsha, Hunan 410205, China; Email: llc@mail.stu.hnfnu.edu.cn

preprocessed images with pixel projection, then located and recognized the target through deep convolution neural network (DCNN), and found through experiment that the method was effective and could recognize the type of workpiece quickly. Binocular vision robot has a wide range of applications in the industrial field and has obvious advantages in modern and automatic production. Its position positioning is also an important and difficult problem. Therefore, the research on its positioning has very important practical values. At present, the accuracy of many methods can not meet the requirements of robot work. To realize the target feature extraction and positioning of binocular vision robot better, this study mainly analyzed the machine learning algorithm, designed a method that extracted features using speeded-up robust features (SURF) and positioned target using back propagation neural network (BPNN), carried out simulation experiments. The present study provides some theoretical support for the application of the method in practice and also makes some contributions to the better application of robots.

2 Positioning principle of binocular vision

Vision is a key technology of robots, which can identify and locate the target. According to the number of cameras, one or two, vision can be divided into monocular and binocular. Binocular vision has higher positioning accuracy and can obtain the three-dimensional information of the target, so it has a more extensive application [12]. The principle of binocular vision is similar to that of human eyes. As there is a distance between the left and right eyes of human, when observing an object, the object will project different positions on the left and right retinas. The deviation of the position is parallax. Binocular vision observes the target through two cameras [14]. According to the parallax of the target, the position of the target can be calculated. When two cameras observe target A at the same time, two coordinates,

$$A_{left} = \left(x_{left}, \ y_{left} \right), \tag{1}$$

$$A_{right} = \left(x_{right}, \ y_{right} \right), \tag{2}$$

are obtained.

According to the trigonometric relations, there is:

$$\begin{cases} x_{left} = f \frac{x_c}{z_c} \\ x_{right} = f \frac{(x_c - B)}{z_c} \\ y = f \frac{y_c}{z_c}, \end{cases} \tag{3}$$

where B refers to the baseline, the distance between the center points of two cameras. If the binocular vision parallax

$$D = x_{left} - x_{right}, \tag{4}$$

then the camera coordinates of A can be expressed as:

$$\begin{cases} x_c = \frac{B x_{left}}{D} \\ y_c = \frac{B y}{D} \\ z_c = \frac{B f}{d}. \end{cases} \tag{5}$$

If parallax D, baseline B and focal length f are determined, the coordinates of the target can be obtained.

3 Target feature extraction

In the two images collected by binocular vision, for location, it is necessary to extract the target area.

The commonly used methods include Scale Invariant Feature Transform (SIFT) algorithm and SURF algorithm. SIFT algorithm has a good accuracy in image feature extraction. SIFT algorithm can extract features well for zoomed and rotated images, but it is difficult to be applied in practice because of its poor real-time performance. SURF algorithm overcomes the above defect and has faster feature extraction speed. Therefore, this study extracted target features using SURF algorithm based on the texture features of the image. SURF is an improvement of Scale Invariant Feature Transform (SIFT) algorithm, which has a high feature extraction speed and can detect stable feature points; therefore it has a good performance in feature extraction [15].

For image I, one point is set as

$$X = (x,\ y)^T, \tag{6}$$

then Hessian matrix $H(X,\ \sigma)$ of scale σ at X can be expressed as:

$$H(x,\ \sigma) = \begin{pmatrix} L_{xx}(x,\ \sigma) & L_{xy}(x,\ \sigma) \\ L_{xy}(x,\ \sigma) & L_{yy}(x,\ \sigma) \end{pmatrix}, \tag{7}$$

where $L_{xx}(x,\ \sigma)$ stands for the convolution of Gaussian second-order partial derivative $\frac{\partial^2}{\partial_x^2}$ with image I at X.

Box filter is used. L_{xx}, L_{xy}, L_{yy} is replaced by D_{xx}, D_{xy}, D_{yy} to obtain the determinant of approximate matrix H_{approx}:

$$\det(H_{approx}) = D_{xx}D_{yy} - (\overline{\omega}D_{xy})^2, \tag{8}$$

where $\overline{\omega}$ stands for a regulation parameter.

SURF algorithm builds image pyramid using the indirect method, divides the scale space according to groups, calculates the image extreme points in each layer of pyramid, sets the threshold value, and performs non-maximal suppression in 3×3×3 neighbourhood on the obtained feature point. Only the point which is larger than the point in the neighbourhood is the feature point.

In order to ensure the invariance of the feature vector, a circular region is divided by taking the feature point as the center and 6 s (s is the scale of the feature point) as the radius. Then the Haar wavelet response operation is performed within the region to calculate the direction corresponding to the maximum Harr response accumulation value, *i.e.*, the main direction of the feature point. Then, taking the point as the center, a square area whose side length is 20 s is selected and divided into 16 subregions. Each region is sampled by 5 s × 5 s, and the Haar wavelet response is calculated. Then a four-dimensional vector is obtained:

$$v = \left(\sum d_x,\ \sum |d_x|,\ \sum d_y,\ \sum |d_y| \right). \tag{9}$$

Sixty-four descriptors can be obtained by combining the vectors in the 16 subregions.

Finally, the steps of the SURF based feature extraction method are as follows. Firstly, the sets of the feature point pairs of the template and the left and right images, I_{left} and I_{right}, are obtained through SURF. Then the feature points which are matched with the template are searched and put into set S_{left} and S_{right}. For the extracted points, the corresponding Euclidean distance is calculated. When the ratio of the two minimum Euclidean distances is smaller than the threshold value, the matching of feature points is successful.

4 Machine learning location algorithm

Machine learning algorithms include decision tree algorithm, Bayesian algorithm, support vector machine (SVM), etc. In this study, the target was positioned using the neural network method from machine learning algorithm. Neural network which is the simulation of human brain has better performance in solving complex nonlinear problems compared to other methods. Moreover, neural network has good adaptivity and fault tolerance. Neural network can be used in positioning. BPNN [16] is the most widely used one, and its structure is shown in Figure 1.

When the target is positioned by BPNN, the image is collected by binocular vision, and the feature points of the target are obtained by feature extraction. The pixel coordinates of the target feature points are taken as
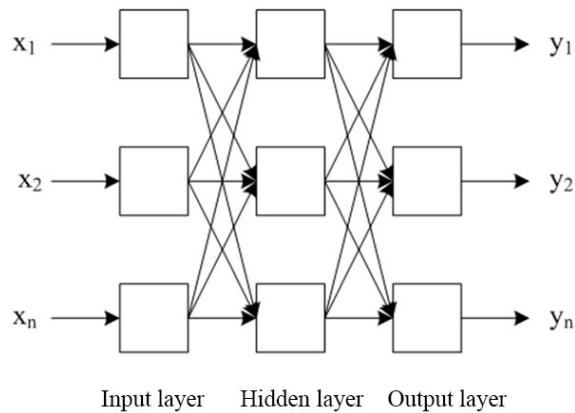
Figure 1: BPNN structure

the input of BPNN, and the world coordinates of the feature points are the output. The specific steps are as follows.

(1) The input vector of BPNN is set as:

$$X = (x_1\ x_2, \ldots,\ x_l), \tag{10}$$

where l is the number of nodes The weight is set as w_{ij}. The input vector of the hidden layer is set as:

$$S = (s_1\ s_2, \ldots,\ s_m), \tag{11}$$

where m stands for the number of neurons, and the output vector was set as:

$$Y = (y_1\ y_2, \ldots,\ y_n), \tag{12}$$

Then:

$$S_j = \sum_{i=0}^{l} w_{ij} x_i, \tag{13}$$

$$y_j = f\left(\sum_{i=0}^{l} w_{ij} x_j\right), \quad i = 1, 2, \cdots, l, \quad j = 1, 2, \cdots m. \tag{14}$$

(2) The input vector of the output layer is set as:

$$R = (r_1\ r_2, \ldots,\ r_n), \tag{15}$$

the output vector is set as:

$$Z = (z_1\ z_2, \ldots,\ z_n), \tag{16}$$

where n stands for the number of nodes, and the weight is set as v_{jk}. Then:

$$r_k = \sum_{j=0}^{m} v_{jk} y_j r, \tag{17}$$

$$z_k = f(r_k) = f\left(\sum_{j=0}^{m} v_{jk},\ v_j\right), \quad k = 1, 2, \cdots, n. \tag{18}$$

(3) If the expected output vector of BPNN is:

$$Q = (q_1, \ q_2, \dots, q_i), \tag{19}$$

then the error between the expected output vector and actual output vector Z is error signal E:

$$E = \frac{1}{2}\sum_{k=1}^{n}(q_k - z_k)^2 \tag{20}$$

After expansion, there is:

$$E = \frac{1}{2}\sum_{k=1}^{n}\left\{ q_k - f\left[\sum_{j=0}^{m}v_{jk}f\left(\sum_{i=0}^{l}w_{ij}x_j\right)\right]\right\}^2 \tag{21}$$

Corrections Δw_{ij} and Δv_{jk} of the weight are:

$$\Delta w_{jk} = \eta\,(q_k - z_k)f\left(\sum_{j=1}^{m}w_{jk}z_jk\right), \tag{22}$$

$$\Delta w_{ij} = \eta\,(q_k - z_k)f\left(\sum_{j=1}^{m}w_{jk}z_jk\right)w_{jk}, \tag{23}$$

where η stands for the learning coefficient.

(4) When the error signal meets the requirements, the calculation stops and the result is output, *i.e.*, the world coordinates of the target feature points.

5 Simulation experiment

The target positioning of binocular vision robot was tested using the SURF and BPNN method proposed in this study. The target was photographed by two charge coupled device (CCD) cameras, and then programming was performed using C++ language in the environment of Visual Studio 2008. Firstly, the feature extraction method was analyzed. Taking an image collected by the camera as an example (Figure 4), the time required for feature extraction of the algorithm was simulated under the condition of no noise and noise multiple of 0.01, 0.03, 0.05 and 0.07 respectively. The results are shown in Figure 2.
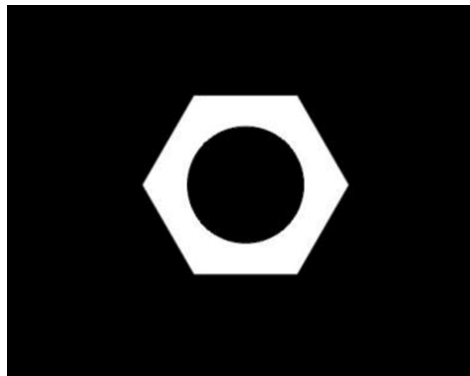

Figure 2: The image of the target

It was seen from Figure 3 that SURF had the highest feature extraction speed in the case of no noise, taking only 0.245 s; with the increase of noise multiple, the calculation time of the algorithm increased gradually, but
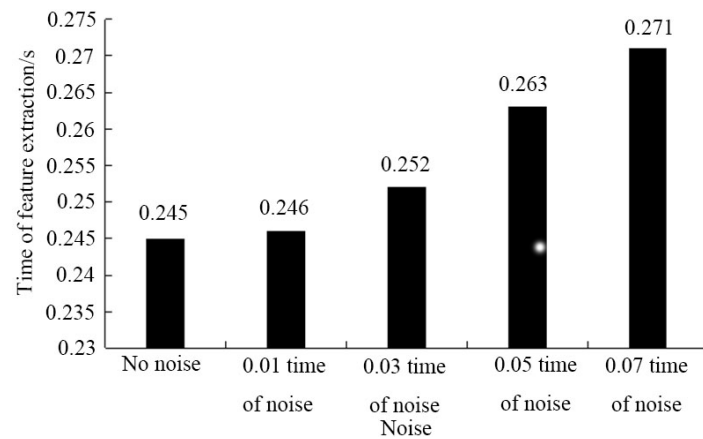
Figure 3: Feature extraction time under different noise conditions

the increase amplitude was very small; when the noise reached 0.07 times, the calculation time was 0.271 s, which was only 0.026s longer than that in the case of no noise, indicating that the feature extraction algorithm designed in this study could meet the real-time requirement.

Two hundred groups of sample images were collected, and the pixel coordinates of the corresponding feature points were extracted. Then 200 groups of sample data were obtained. One hundred and fifty groups of data were used for BPNN training, and the remaining 50 groups of data were used for positioning test. Pixel coordinates p_l, q_l, p_r and q_r of feature points of left and right images obtained by binocular vision were the input of BPNN, and world coordinates x_w, y_w and z_w of the feature points were the output of BPNN. The BPNN method used in this study was compared with the SVM method [17]. The testing results are shown in Table 1.

Table 1: Positioning results

No.	Expected output			Actual output					
				SVM			BPNN		
	x_w	y_w	z_w	x_w	y_w	z_w	x_w	y_w	z_w
1	271	110	0	273.1625	110.918	0.0435	271.3648	110.0254	0.0000
2	289	133	0	291.1452	134.4194	0.0546	290.3311	133.5268	0.0001
3	310	112	0	311.9584	113.303	0.0514	310.2587	111.5896	0.0000
4	456	128	0	457.8526	129.5284	0.0425	455.9658	128.6358	0.0000
5	359	136	0	361.0215	137.6771	0.0442	358.1258	136.7845	0.0001
6	224	127	0	225.4962	128.7792	0.0454	225.2615	127.8866	0.0001
7	268	109	0	270.8219	109.9513	0.0485	268.1258	109.0587	0.0001
8	316	129	0	318.4733	130.4213	0.0436	315.6258	129.5287	0.0000
9	478	164	0	481.2193	166.578	0.0465	478.5264	165.6854	0.0000
10	532	151	0	535.5889	152.3678	0.0485	531.6589	150.5248	0.0000
11	276	154	0	278.5819	155.278	0.0484	277.3256	154.3854	0.0001
12	289	167	0	290.8136	169.1523	0.0436	288.1028	168.2597	0.0000
......	......	......	......	......	......	......	......	......	......
50	315	156	0	316.4565	157.1201	0.0462	315.9215	155.6914	0.0001

It was seen from Table 1 that the output result of SVM was slightly different with the expected output, but the actual output of BPNN was very close to the expected output, and the error of the output of z_w especially was almost 0. The error of every time of positioning of the two methods was counted, and the results are shown in Figures 4 and 5.
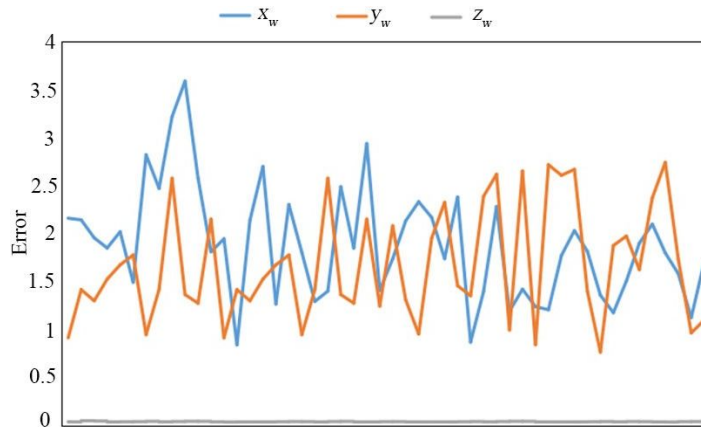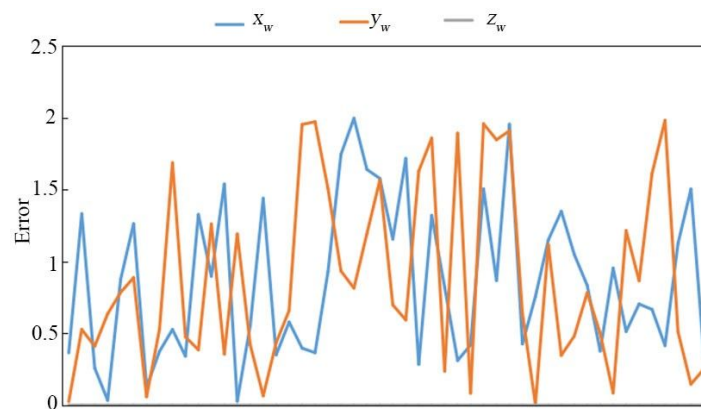
Figure 4: Error changes of SVM

Figure 5: Error changes of BPNN

After calculation, it was found that the maximum error values of x_w, y_w, and z_w in SVM were 3.5889, 2.7437, and 0.0546 respectively, and the minimum errors were 0.8732, 0.7694, and 0.411 respectively; in BPNN, the maximum errors of x_w, y_w, and z_w were 1.9954, 1.9812, and 0.0001 respectively, and the minimum errors were 0.0280, 0.0069 and 0.0000 respectively. The above results showed that the error of SVM was significantly larger than BPNN, indicating that the positioning precision of BPNN was higher. It was found that the error of BPNN in X and Y directions was slightly larger, and the error was almost 0 in the Z direction. The changes of error showed that the error was always smaller than 2, which could meet the positioning requirements of robots.

6 Discussion

Machine learning covers many aspects of knowledge, such as probability theory, statistics, etc., which is a very important research direction in artificial intelligence [18]. It has been widely used in solving complex engineering and scientific problems, such as natural language processing [19], pattern recognition [20], biological information processing [21], machine vision [22], etc. The algorithms of machine learning include decision tree [23], random forest [24], Bayesian [25], etc. In this study, the target positioning method was designed using BPNN method.

Before a robot positions a target, feature extraction of the target is needed. In this study, the feature extraction of the target was achieved by SURF method. Then the pixel coordinates of the extracted feature

points were used as the input of BPNN to locate the target. The experimental results showed that SURF method had a good performance in target feature extraction as it completed feature extraction in a short time and was less affected by noise. It was seen from Figure 3 that the increase amplitude of the calculation time of the algorithm was very small, around 0.2 s, though there was an increase with the increase of noise. The positioning experiment showed that the positioning precision of BPNN was significantly superior to that of SVM, the output position coordinates of BPNN were very close to the actual coordinates, and the errors in three directions were very small, which could meet the needs of robots well in the actual work.

This study found that the SURF method and BPNN model showed good performance and had strong applicability in solving the positioning problem of robots; however, there are some limitations that need to be improved in future works:

(1) comparative study was not carried out on more machine learning methods;
(2) the BPNN method was not further optimized to enhance the positioning precision.

7 Conclusion

In order to solve the problem of robot target location, the target features were extracted using the SURF method, and then the location of the target was realized by the BPNN model. It was found from the experiment that:

(1) the SURF method was less interfered by noise and had a high extraction speed, showing a good performance in the target feature extraction;
(2) the positioning result of the BPNN method was superior to that of SVM and had very small errors, and the error in X and Y directions was slightly large, but not more than 2.

The experimental results verify the effectiveness of the proposed method for target feature extraction and location, which can be promoted and applied in practice.

References

[1] E. Clotet, D. Martínez, J. Moreno, M. Tresanchez and J. Palacín, Development of a High Mobility Assistant Personal Robot for Home Operation, Adv. Intell. Syst. Comput. **376** (2015), 65-73.

[2] R. J. Guo and J. S. Zhao, Topological principle of strengthened connecting frames in the stretchable arm of an industry coating robot, Mech. Mach. Theory **114** (2017), 38-59.

[3] S. Erfani, A. Jafari and A. Hajiahmad, Comparison of two data fusion methods for localization of wheeled mobile robot in farm conditions, Artif. Intell. Agric. **1** (2019), 48-55.

[4] K. Zhang, X. Huang, Y. Gao, W. Liang, H. Xi, et al., Robot-Assisted Versus Laparoscopy-Assisted Proximal Gastrectomy for Early Gastric Cancer in the Upper Location, Cancer Control **25** (2018), 107327481876599.

[5] S. Y. Choi and J. H. Yang, Analysis of the Human Performance and Communication Effects on the Operator Tasks of Military Robot Vehicles by Using Extended Petri Nets, Korean J. Comput. Design Eng. **22** (2017), 162-171.

[6] H. K. Kim, H. S. Sim and W. J. Hwang, A Study on a Path Planning and Real-Time Trajectory Control of Autonomous Travelling Robot for Unmanned FA, J. Korean Soc. Ind. Converg. **19** (2016), 75-80.

[7] V. Belle and H. J. Levesque, Robot location estimation in the situation calculus, J. Appl. Logic. **13** (2015), 397-413.

[8] L. Zhang, H. Liu, C. Luo, G. Bian and W. Wu, Target recognition of indoor trolley for humanoid robot based on piecewise fitting method, Int. J. Adapt. Control **33** (2019), 1319-1327.

[9] L. Luo, X. Zou, J. Xiong, Y. Zhang, H. Peng, et al., Automatic positioning for picking point of grape picking robot in natural environment, Trans. Chin. Soc. Agric. Eng. **31** (2015), 14-21(8).

[10] L. Yan and D. Xiong, Mobile motion robot indoor passive RFID location research, Int. J. Rf. Tech. Res. Appl. **9** (2018), 113-129.

[11] X. Wu, X. Ling and J. Liu, Location Recognition Algorithm for Vision-Based Industrial Sorting Robot via Deep Learning, Int. J. Pattern Recogn. **33** (2019), 1955009.1-1955009.18.

[12] P. Ueareeworakul and S. Saiyod,

[13] 2017 9th International Conference on Knowledge and Smart Technology (KST) - Obstacle detection algorithm for unmanned aerial vehicles using binocular stereoscopic vision. Int. Confon. Knowledge & Smart Technology (2017), 332-337.

[14] J. Turski, On binocular vision: The geometric horopter and Cyclopean eye, Vision Res. **119** (2016), 73-81.

[15] G. Dan, M. A. Khan and V. Fodor, Characterization of SURF and BRISK Interest Point Distribution for Distributed Feature Extraction in Visual Sensor Networks, IEEE T. Multimedia **17** (2015), 591-602.

[16] J. Fan, J. Zhong, J. Zhao and Y. Zhu, BP neural network tuned PID controller for position tracking of a pneumatic artificial muscle, Tech. Health Care: Offic. J. Eur. Soc. Eng. Med. **23** (2015), S231-S238.

[17] D. Chen, L. Wang, and L. Li, Position computation models for high-speed train based on support vector machine approach, Appl. Soft Comput. **30** (2015), 758-766.

[18] M. I. Jordan and T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science **349** (2015), 255-260.

[19] Y. Ni, D. Barzman, A. Bachtel, M. F. Griffey, A. Osborn, et al., Finding Warning Markers: Leveraging Natural Language Processing and Machine Learning Technologies to Detect Risk of School Violence, Int. J. Med. Inform. **139** (2020), 104137.

[20] D. S. Bulgarevich, S. Tsukamoto, T. Kasuya, M. Demura and M. Watanabe, Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures, Sci. Rep. **8** (2018), 2078.

[21] N. Kriegeskorte, Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing, Ann. Rev. Vis. Sci. **1** (2015), 417-446.

[22] R. Shanmugamani, M. Sadique and B. Ramamoorthy, Detection and classification of surface defects of gun barrels using computer vision and machine learning, Measurement **60** (2015), 222-230.

[23] Y. M. Shi, W. W. Chen and Y. F. Zhu, Study on Prediction Model of Number of Rainstorm Days in Summer Based on C5.0 Decision Tree Algorithm, Meteorol. Environ. Res. **10** (2019), 60-64.

[24] M. G. Appley, S. Beyramysoltan and R. A. Musah, Random Forest Processing of Direct Analysis in Real-Time Mass Spectrometric Data Enables Species Identification of Psychoactive Plants from Their Headspace Chemical Signatures, Acs Omega **4** (2019), 15636-15644.

[25] N. Bin, J. W. Wu and F. Hu, Spam Message Classification Based on the Naieve Bayes Classification Algorithm, IAENG Int. J. Comput. Sci. **46** (2019), 46-53.